L-1 IDENTITY SOLUTIONS™

BIOMETRICS DIVISION

Biometric Identification: Metrics & Models

Brian Martin

March 2010

Overview

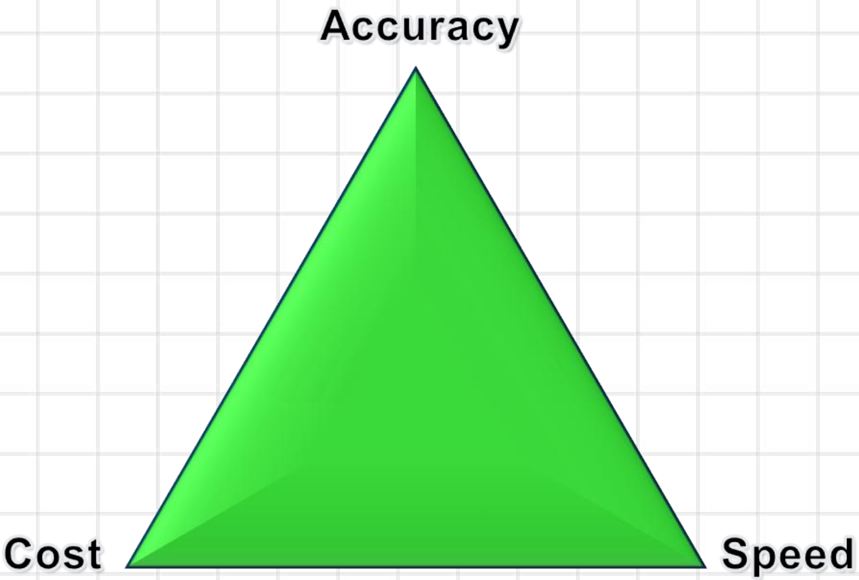
- **Back of envelope approach to biometric ID systems – and in particular the Search Engine Backend:**
  - **What are (some of) the relevant metrics?**
  - **Can we use what we measure to model performance? How?**

- **If interested in a more precise examination of the topic (free dl):**
  - **"National Biometric Test Center Collected Works 1997-2000"**
    Dated, but still many relevant points for 1:N search systems
    - Wayman, 2000
    - http://www.engr.sjsu.edu/biometrics/publications.html
  - **"Matching Performance for the US-VISIT IDENT System Using Flat Fingerprints (NIST IR-7110)"**
    One of few published results on large scale testing
    - Wilson, Garris, & Watson, 2004
    - ftp://sequoyah.nist.gov/pub/nist_internal_reports/ir_7110.pdf

Metrics

Biometric Search Engine Metrics

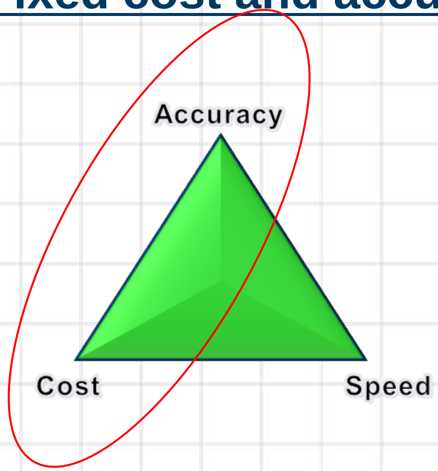
- There is an analogy between biometric ID search engine trade-offs and the project management triangle:
  “Good, Fast, Cheap… Pick any two”



- There are arguably several other dimensions… these seem to capture the general concepts of most
- All axes are interrelated…

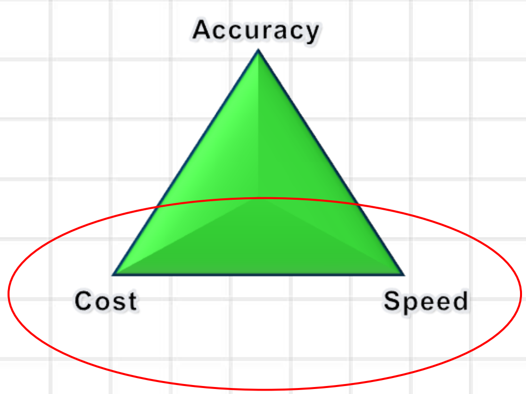
Biometric Search Engine Metrics

Fixed cost and accuracy



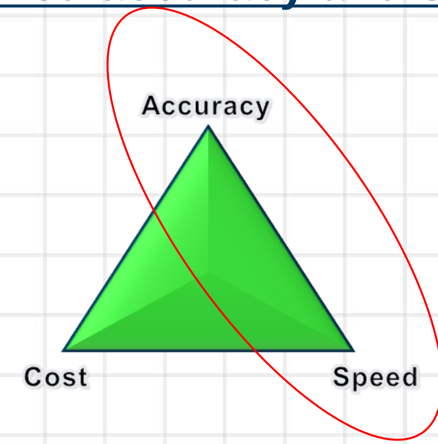
-Hardware constrained solution such as access control.
-Speed varies with users, database size, etc…

Fixed cost and speed



-Accuracy is determined by investment cost and time
- Perhaps MBGC is an example. Cost and speed are determined by participant.

Fixed accuracy and speed



- RFP sets requirements for speed and accuracy
- Cost is the dependent variable

Biometric Search Engine Tests

- **There are many types of biometric system performance tests. Some of the most common:**
  1. **Document and explore current state of the art**
     - Only test accuracy (“MBGC”, PFT, etc…)
     - Sometimes accuracy and speed (IREX, ELFT, FVC)
     - Helps answer, “Theoretically, can biometrics provide a solution to a particular problem?”
  2. **Validate existing system performance**
     - Accuracy, speed and cost are considered
  3. **Collect data that suggests future system performance (procurements)**
     - Will a system meet requirements at a smaller scale?
     - How much will the final, larger, system cost?
     - Accuracy, speed, cost, and a model for scaling

Metrics: Time

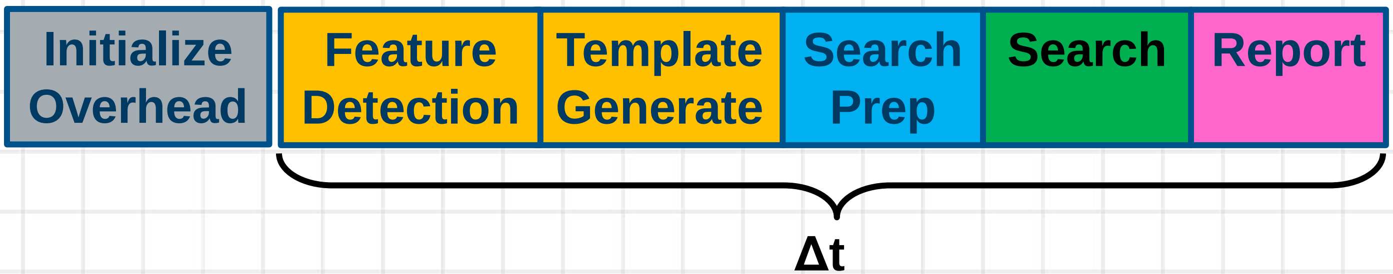
Metric: Time

- **Processing time (enroll and match) depends on**
  - **Data quality characteristics**
  - **Imposter / Genuine (strength of match)**
  - **Database size**
  - **HW dependencies**
    - CPU bound
    - Memory bound
    - Instruction set support (SSE, NUMA, etc…)
    - Scaling approach (multi-core, system architecture, etc…)

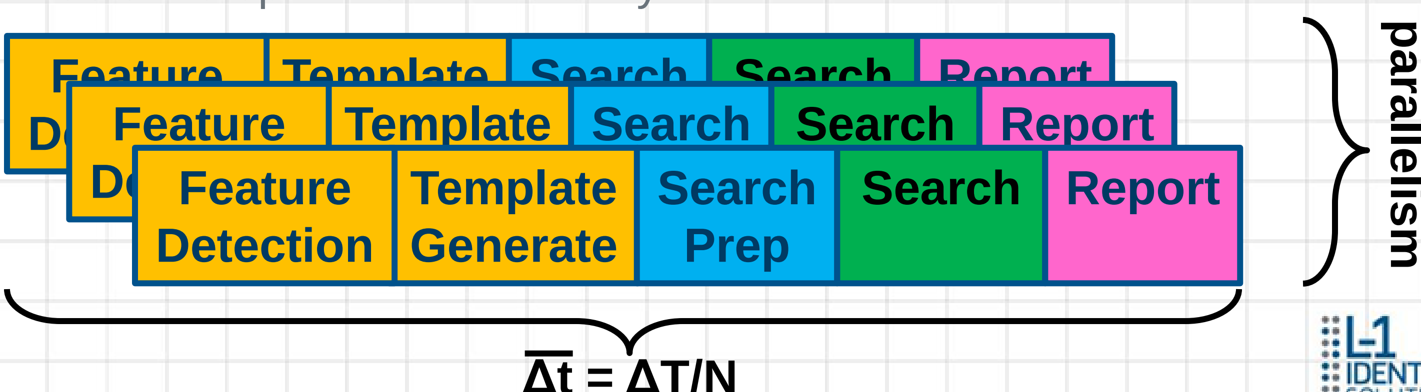
L-1 IDENTITY SOLUTIONS™ BIOMETRICS DIVISION

Metric: Time

- **Two basic flavors of the time metric:**
  - **Latency – how long**
    - Real time metric



  - **Throughput – how much (average rate)**
    - Batch processing
    - Easier to optimize than latency



Metric: Time

- **Measure**
  - Latency and throughput
- **Should document**
  - Database size
  - Data quality aspects
  - Imposter / Genuine distribution
  - Hardware description
  - One-time overhead, measures at different gallery sizes
  - Architecture (multi-core, multi-server) overhead

Metrics: Accuracy

L-1 IDENTITY SOLUTIONS™ BIOMETRICS DIVISION

Metric: Accuracy

- **Accuracy depends on**
  - **Algorithmic sophistication**
    - Feature detection
    - Feature matching
  - **Biometric sample quality**
- **Most independent tests do an excellent job measuring accuracy on specific database samples**
  - **Most tests become dated quickly**

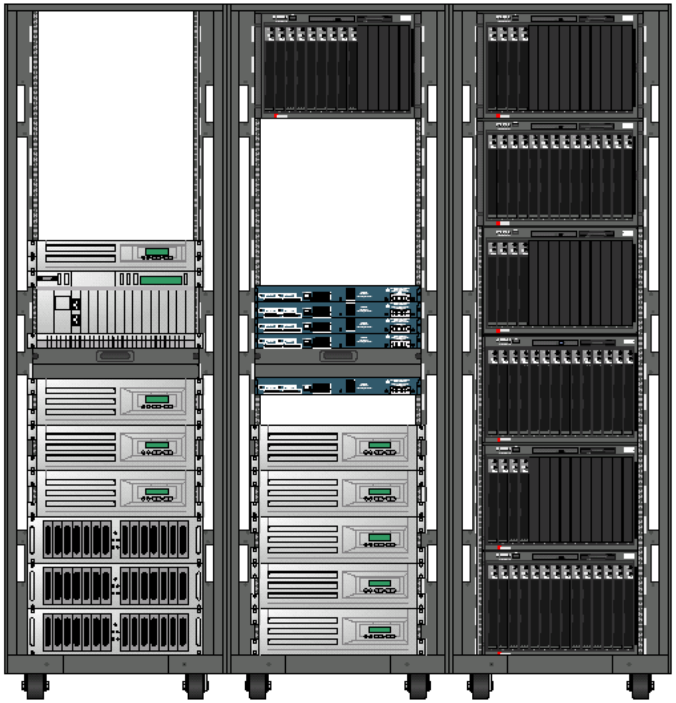
Metric: Accuracy

- **Measure for 1:1**
  - **Tradeoff between FMR and FNMR (aka FAR and FRR)**
    - ROC or DET curve
  - **FTE**
  - **Examples: IREX, MINEX, PFT, FVC**
- **Measure for 1:N**
  - **Tradeoff between FPIR and FNIR (aka FAR and FRR, FMR and FNMR, Selectivity, 'Alarm' rates)**
    - Use open set
    - Measures depends on result list size. Here we assume list size of 1.
    - Names of the metrics seem to drift from document to document – why?
    - $FNIR \sim FNMR_{1:1}$
    - $FPIR \sim FMR_{1:1} \times N_{DB}$
  - **FTE**
  - **CMC (hit rate) useful when every search is reviewed by human (latent)**
  - **Examples: (FpVTE, FRVT, ICE), ELFT**
- **Also should document the FTA rate if possible**

Metrics: Cost

Metric: Cost

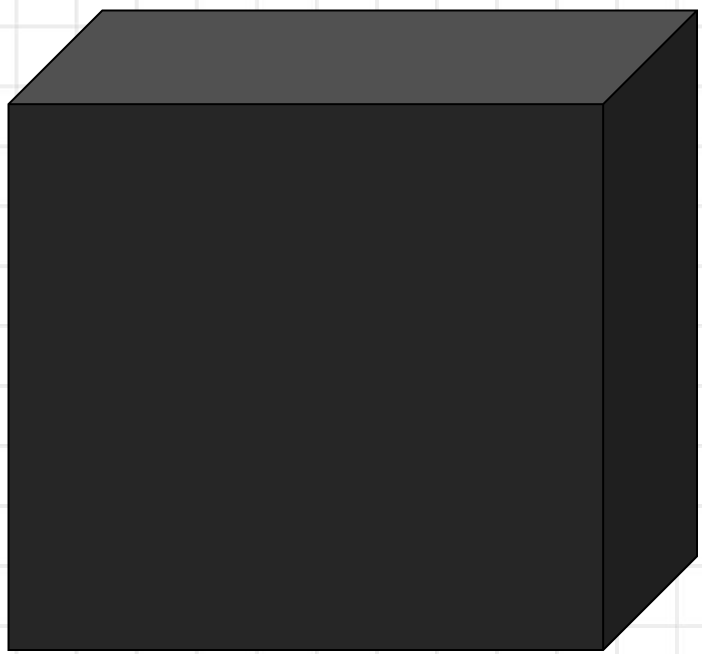
- **The cost metric is a direct reflection of**
  - **Hardware required for the solution = <u>system footprint</u>**
    - Depends on computational efficiency (speed, size)
    - Engineering sophistication of matcher
    - System architecture
  - **Maintainability (Power, Cooling, Support, etc…)**
  - **Human review workload (accuracy dependent)**
- **Cost usually not an independent variable in testing.**
- **Cost can be reasonably estimated if the system architecture and other metrics are understood**
  - **Modeling cost can be non-linear with project scale (both ways!)**



ID System Modeling

ID System Models

- **Given test results, how do we use measurements?**
  - Model accuracy, speed, and cost for some other system
- **Can we keep things simple?**



ID System Models – one problem

- **Real world 'black-box'**
  - Meets the requirements
  - Larger databases
  - Minimize cost = efficient

- **Lab test 'black-box'**
  - Configured for best accuracy
  - Smaller databases
  - Maximize cost/time to limits

ID System Models

- Lab testing can be useful, but the difference between the black-box in the lab and the black-box AFIS search engine must be acknowledged
  - Measuring more than accuracy helps bring this to light

- Now, lets look at a couple models for biometric search engines and examine how we can use test results

L-1 IDENTITY SOLUTIONS™ BIOMETRICS DIVISION

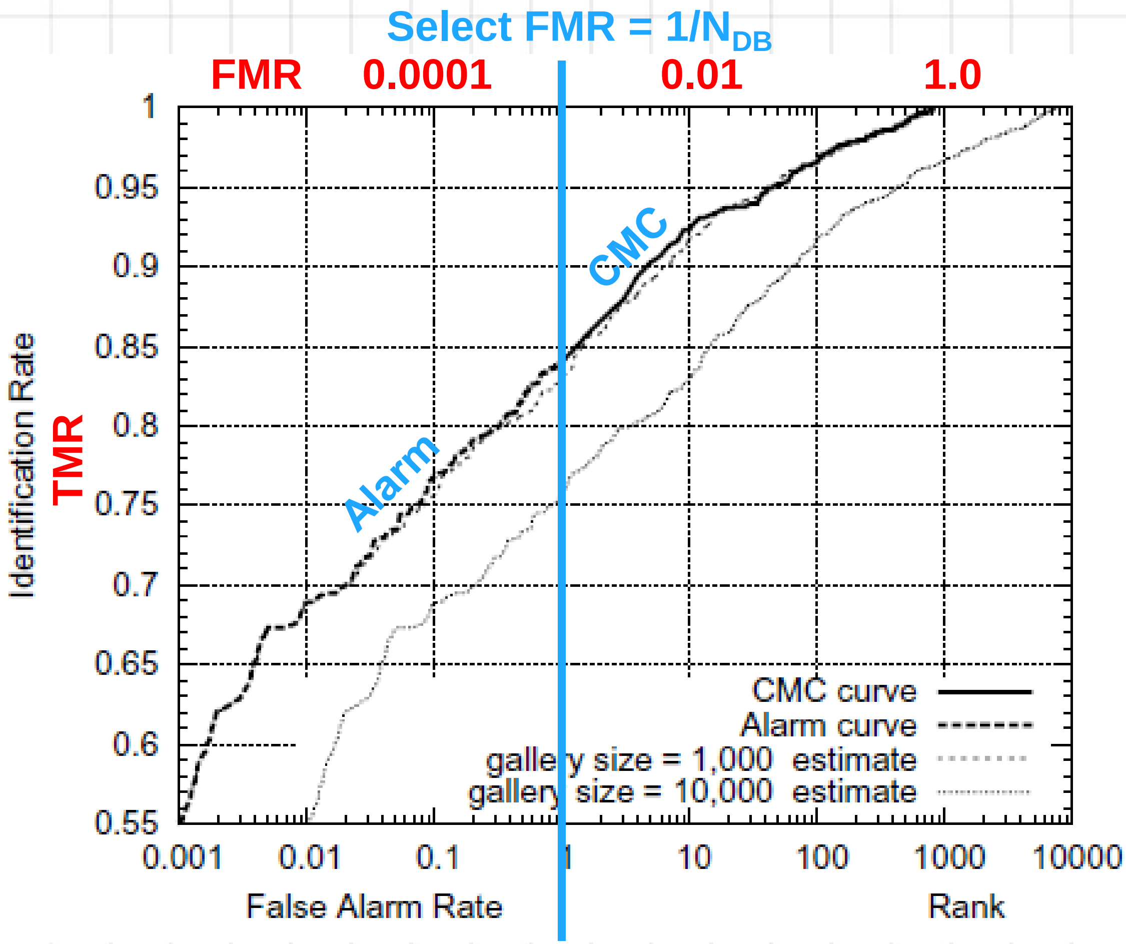
Model: 1-to-N = N x (1-to-1)

- The most common assumption for biometric identification systems is that 1:N = N x (1:1)
  - Rarely the case in practice
  - Easy to model and can provides good 'back of envelope' estimations for simple (small) ID systems
- Latency ≈ N x 1:1 latency
- Throughput ≈ 1/Latency
- System size estimated by:
  - Number of CPU cores directly calculated from throughput or latency
  - Amount of RAM required calculated from template size
- Accuracy modeling previously presented (BSYM06)

L-1 IDENTITY SOLUTIONS™ BIOMETRICS DIVISION

Model: 1-to-N = N x (1-to-1)

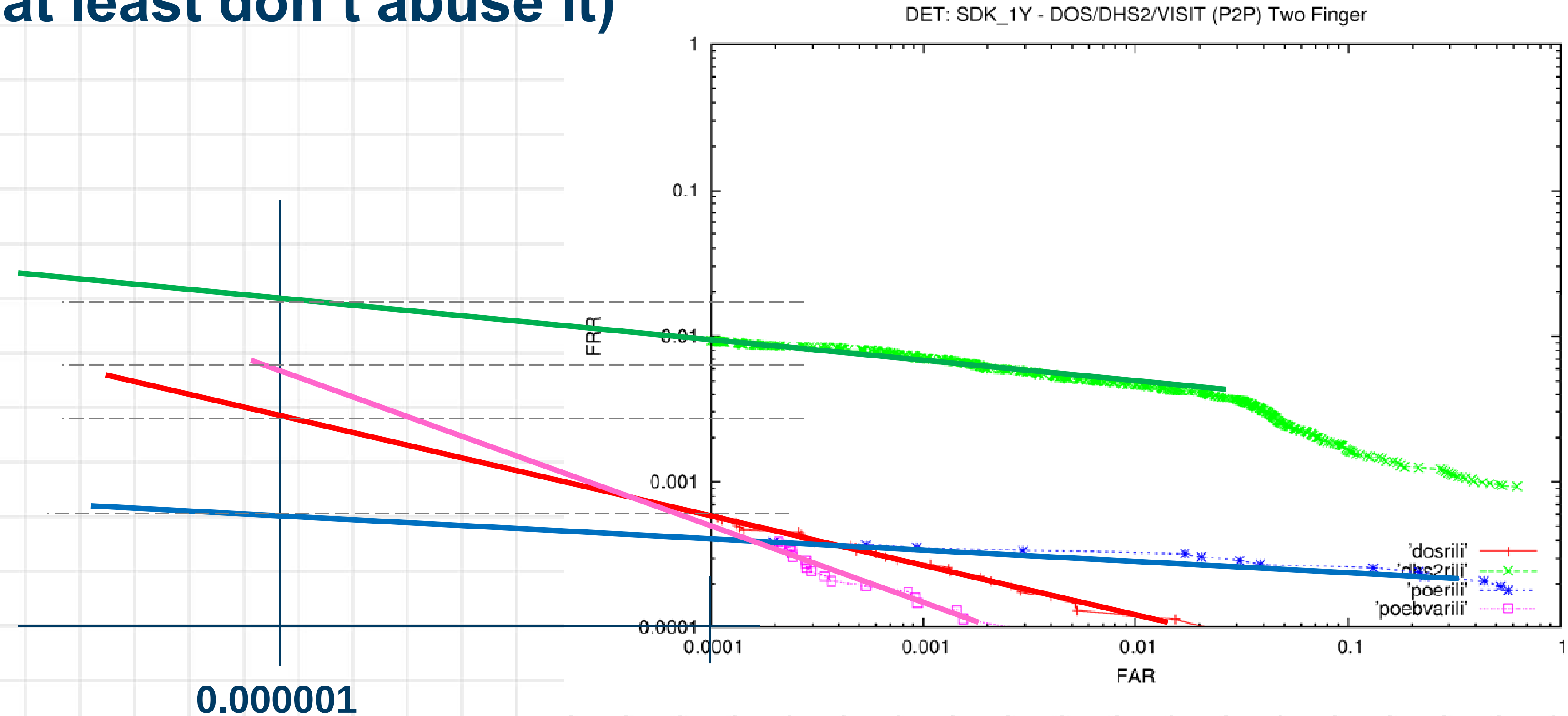
- **The 1:1 ROC curve can be used to estimate the CMC and alarm curves**



Model: 1-to-N = N x (1-to-1)

- **Avoid this very tempting technique… !**

(at least don't abuse it)

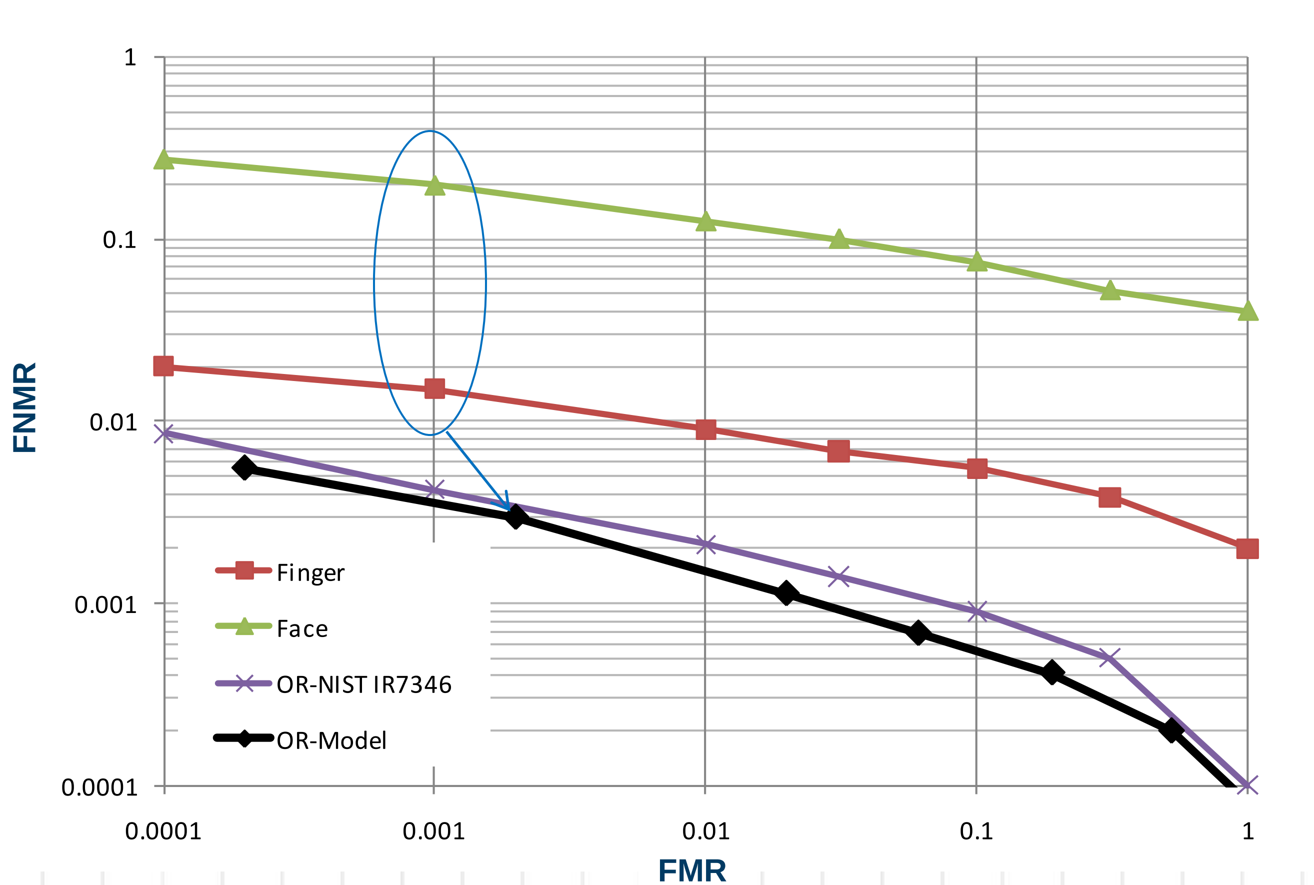


Model: 1-to-N = N x (1-to-1) FUSION

- **The ROC can also be used to model multi-modal fusion**
  - Needed: 2 ROC curves of uncorrelated biometrics and a calculator
- **This result of OR rule fusion is easy to derive**
  - At each operating point, there are 2 scores S_1 and S_2
  - Chance the person is rejected is the probability of both scores being below the threshold (sum the probabilities of independent events)
    - $FNMR_{1|2}=(FNMR_1)x(FNMR_2)$
  - The person is falsely matched when S_1 > Tor S_2 >T or when S_1 and S_2 are BOTH > T. This is 1 minus probability that the person is correctly rejected. This happens when BOTH S_1 and S_2 are correctly rejected. Again this is the sum of probabilities of independent events.
    - $FMR_{1|2} = 1 – (TNMR_1)x(TNMR_2) = 1 – (1 – FMR_1)x(1 – FMR_2)$
    - At low FMRs, $FMR_{1|2} \approx FMR_1 + FMR_2$

Model: 1-to-N = N x (1-to-1) FUSION

- **Example from NIST IR-7346 fusion study.**
- **Face and Finger fusion compared to score based result:**

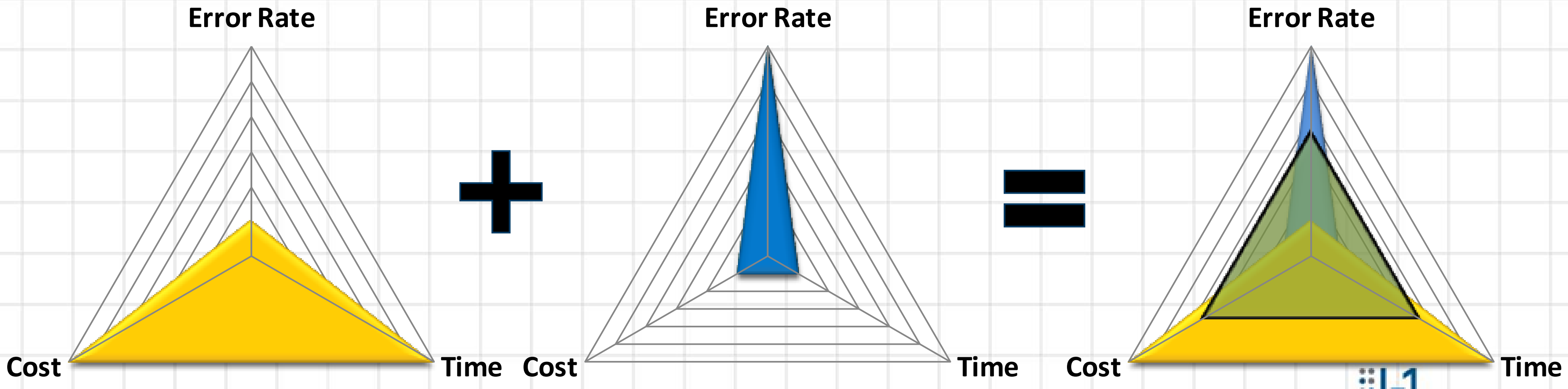


Model: 1-to-N = N x (1-to-1)

- **Where does this model based on 1-to-1 matching breakdown?**
  - **Timing**
    - Scaling behavior not clear
      - One-time latencies
      - Threading efficiency
      - Scoring overhead
  - **Accuracy**
    - When gallery normalization is used
    - When multiple matchers are used selectively
- **The simple model doesn't handle advanced matching approaches which better scale to large DB sizes.**
- **The next model shows why things aren't so simple.**

Model: 1-to-N

- One main issue with the previous model is the assumption that the a single 1:1 match event is repeated for 1:N search. This is typically not true for large scale, or high throughput systems.
- Most modern biometric identification systems employ a multi-stage matching approach for improving speed – this breaks our simple model



Model: 1-to-N

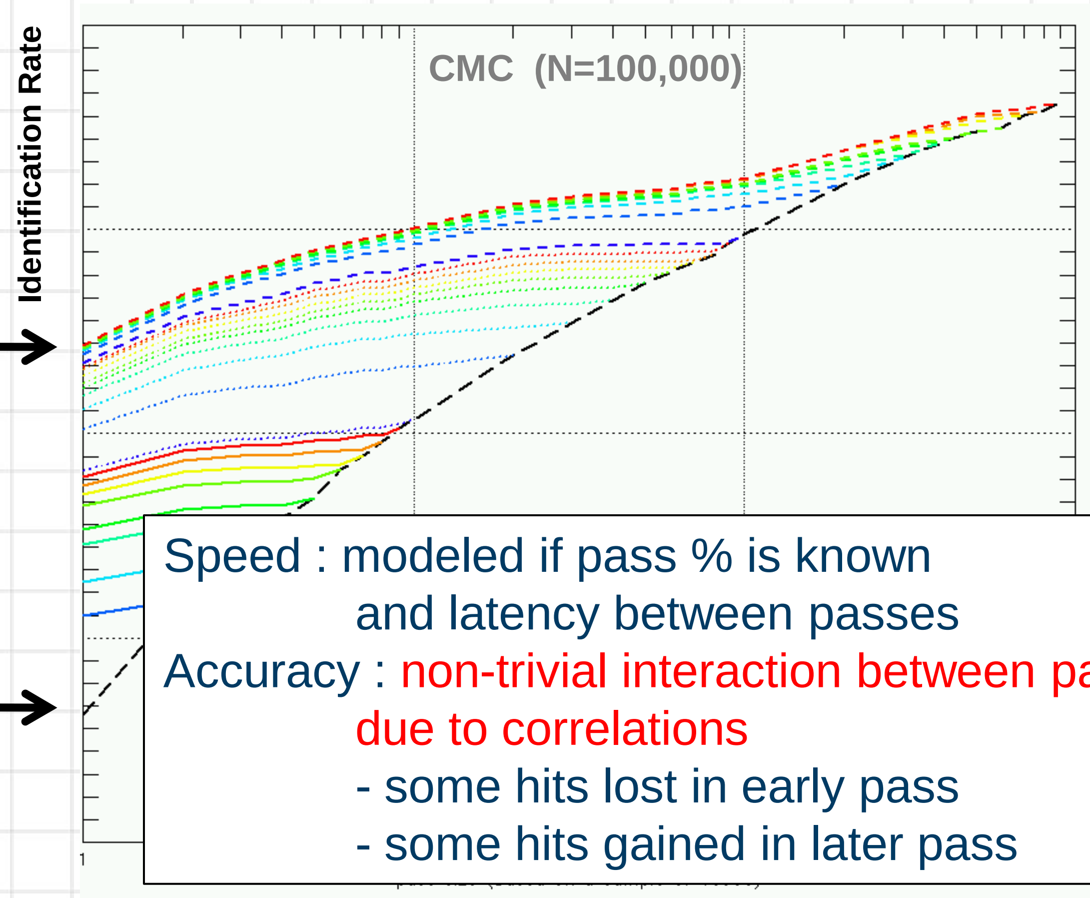
- **Multi-stage matching breaks the Nx1-to-1 model**
- **Two examples of multi-stage matching**
  1. **Incremental**
     - On each match attempt, effort depends on
       - Sample quality
       - Preliminary evaluation of the likelihood of match
  2. **Multi-pass**
     - Rank or filter all matches, apply additional matching effort on most promising candidates
       - Good for combining very different matching approaches including filtering and multi-modal
       - A cascade of the previous Nx(1:1) model

Multi-Pass Matching

- **Two pass face identification system**

Pass 2
Better CMC

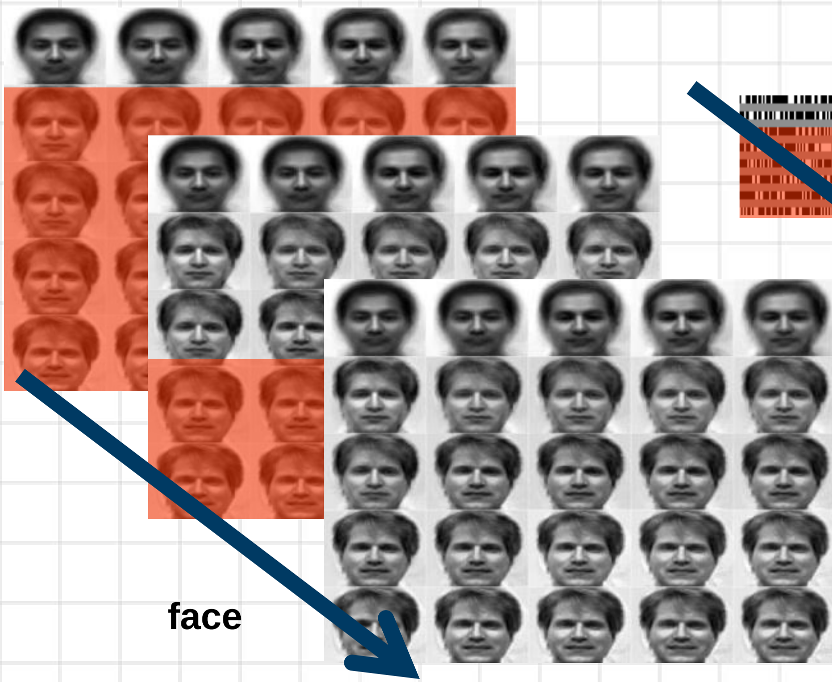
Pass 1

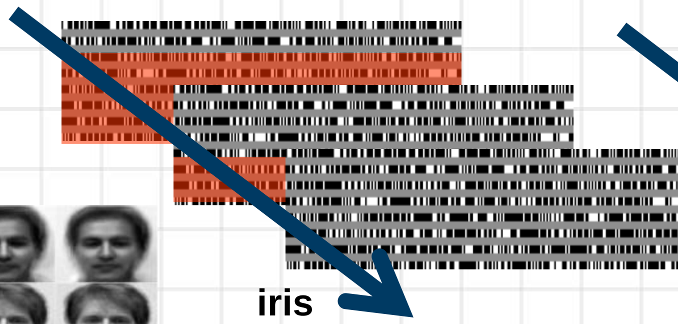


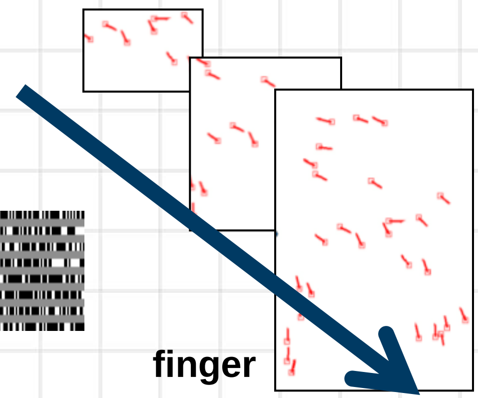
Speed : modeled if pass % is known
and latency between passes
Accuracy : non-trivial interaction between passes
due to correlations
- some hits lost in early pass
- some hits gained in later pass

Incremental Matching

- **Increase intensity of matching as needed**
  - Increment amount of data examined (minutiae, iris code bits, Eigen face coefficients, etc…)
  - Incremental algorithmic intensity







Speed : depends on data
Accuracy : some loss, mostly in
non-relevant operating region,
not trivial to model

Face image from www.cs.princeton.edu/~cdecoro/eigenfaces

L-1 IDENTITY SOLUTIONS
BIOMETRICS DIVISION

Multi-Stage Matching

- **Unfortunately there is no good generalized ‘black-box’ or ‘gray box’ model**
  - Matching speed for large systems not easily predictable
  - Accuracy on different data not easily predictable

- **Therefore, one needs some understanding of how the system works for relatively accurate modeling of larger systems**
  - Also requires empirical measurements for several parts of the system separately from the whole system

Other Considerations to Keep in Mind

- **Search engine architecture matters**
  - **The models presented here do not take into account the workload distribution over several match servers**
    - Synchronizing, collation of results, etc..
    - How does architecture scale matchers?
      - Divide and conquer – per thread, per machine?
      - Parallel search – how many at once? Efficiency?
  - **System overhead can overwhelm individual matcher timing**

Take Home Messages

- **Performance testing of a biometric search engine is multi-faceted, not just accuracy**
- **We can use simple 1:1 measurements to get very rough estimates for small biometric ID systems**
  - In general, everything is much more complicated though
- **Use caution when trying to extending lab testing results to other (larger) system requirements**
  - Please don't assume anything – talk to vendors about how their system scales if needed

Thank You

bmartin@L1ID.com